REQUEST FOR ACCESS OF ABANDONED APPLICATION UNDER 37 CFR 1.14(a)

In re Application of	
Application Number: 08/227,525	Filed
Group Art Unit	Examiner

Paper No. 35

Assistant Commissioner for Patents
Washington, DC 20231

I hereby request access under 37 CFR 1.14(a)(3)(iv) to the application file record of the above-identified ABANDONED application, which is: (CHECK ONE)

__V__ (A) referred to in United States Patent Number 5733729, column 1,

___ (B) referred to in an application that is open to public inspection as set forth in 37 CFR 1.11, i.e., Application No. _____, filed _____, on page _____ of paper number _____,

___ (C) an application that claims the benefit of the filing date of an application that is open to public inspection, i.e., Application No. _____, filed _____, or

___ (D) an application in which the applicant has filed an authorization to lay open the complete application to the public.

Please direct any correspondence concerning this request to the following address:

_____

_____

_____

Shoaib Chayour
Signature

06-19-98
Date

Shoaib Chayour
Typed or printed name

6/19/98

FOR PTO USE ONLY

Approved by: _____ (initials)

Unit: _____

COMPUTER-AIDED PROBABILITY BASE CALLING FOR ARRAYS OF NUCLEIC ACID PROBES ON CHIPS

GOVERNMENT RIGHTS NOTICE

Portions of the material in this specification arose under ie cooperative agreement 70NANB5H1031 between ffymetrix, Inc. and the Department of Commerce through ie National Institute of Standards and Technology.

COPYRIGHT NOTICE

SOFTWARE APPENDIX

A Software Appendix comprising twenty one (21) sheets s included herewith.

BACKGROUND OF THE INVENTION

The present invention relates to the field of computer ystems. More specifically, the present invention relates to omputer systems for evaluating and comparing biological equences.

Devices and computer systems for forming and using rrays of materials on a substrate are known. For example, 'CT application WO92/10588, incorporated herein by ref-rence for all purposes, describes techniques for sequencing r sequence checking nucleic acids and other materials. Arrays for performing these operations may be formed in rrays according to the methods of, for example, the pio-neering techniques disclosed in U.S. Pat. No. 5,143,854 and U.S. patent application Ser. No. 08/249,188, both incorpo-rated herein by reference for all purposes.

According to one aspect of the techniques described herein, an array of nucleic acid probes is fabricated at known locations on a chip or substrate. A fluorescently labeled nucleic acid is then brought into contact with the chip and a scanner generates an image file (also called a cell file) indicating the locations where the labeled nucleic acids bound to the chip. Based upon the image file and identities of the probes at specific locations, it becomes possible to extract information such as the monomer sequence of DNA or RNA. Such systems have been used to form, for example, arrays of DNA that may be used to study and detect mutations relevant to cystic fibrosis, the P53 gene (relevant to certain cancers), HIV, and other genetic characteristics.

Innovative computer-aided techniques for base calling are disclosed in U.S. patent application Ser. No. 08/327,525, which is incorporated by reference for all purposes. However, improved computer systems and methods are still needed to evaluate, analyze, and process the vast amount of information now used and made available by these pioneering technologies.

SUMMARY OF THE INVENTION

An improved computer-aided system for calling unknown bases in sample nucleic acid sequences from multiple nucleic acid probe intensities is disclosed. The present invention is able to call bases with extremely high accuracy (up to 98.5%). At the same time, confidence information may be provided that indicates the likelihood that the base has been called correctly. The methods of the present invention are robust and uniformly optimal regardless of the experimental conditions.

According to one aspect of the invention, a computer system is used to identify an unknown base in a sample nucleic acid sequence by the steps of: inputting a plurality of hybridization probe intensities, each of the probe intensities corresponding to a nucleic acid probe; for each of the plurality of probe intensities, determining a probability that the corresponding nucleic acid probe best hybridizes with the sample nucleic acid sequence; and calling the unknown base according to the nucleic acid probe with the highest associated probability.

According to another aspect of the invention, an unknown base in a sample nucleic acid sequence is called by a base call with the highest probability of correctly calling the unknown base. The unknown base in the sample nucleic acid sequence is identified by the steps of: inputting multiple base calls for the unknown base, each of the base calls having an associated probability which represents a confidence that the unknown base is called correctly; selecting a base call that has a highest associated probability; and calling the unknown base according to the selected base call. The multiple base calls are typically produced from multiple experiments. The multiple experiments may be performed on the same chip utilizing different parameters (e.g., nucleic acid probe length).

According to yet another aspect of the invention, an unknown base in a sample nucleic acid sequence is called according to multiple base calls that collectively have the highest probability of correctly calling the unknown base. The unknown base in the sample nucleic acid sequence is identified by the steps of: inputting multiple probabilities for each possible base for the unknown base, each of the probabilities representing a probability that the unknown base is an associated base; producing a product of probabilities for each possible base, each product being associated with a possible base; and calling the unknown base according to a base associated with a highest product. The multiple base calls are typically produced from multiple experiments. The multiple experiments may be performed on the same chip utilizing different parameters (e.g., nucleic acid probe length).

According to another aspect of the invention, both strands of a DNA molecule are analyzed to increase the accuracy of identifying an unknown base in a sample nucleic acid sequence by the steps of: inputting a first base call for the unknown base, the first base call determined from a first nucleic acid probe that is equivalent to a portion of the sample nucleic acid sequence including the unknown base; inputting a second base call for the unknown base, the second base call determined from a second nucleic acid probe that is complementary to a portion of the sample nucleic acid sequence including the unknown base; selecting one of the first or second nucleic acid probes that has a base at an interrogation position which has a high probability of producing correct base calls; and calling the unknown base according to the selected one of the first or second nucleic acid probes.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates an example of a computer system used to execute the software of the present invention;

40-041 Henm 2

A006